DOCUMENT RESUME

ED 137 411 TM 006 220

AUTHOR Crocker, Linda; Benson, Jeri
TITLE Effects of Examinee Response Changes on Item and Test Characteristics.
SPONS AGENCY Florida Univ., Gainesville. Inst. for Development of Human Resources.
PUB DATE [Apr 77]
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; Complexity Level; Correlation; Graduate Students; *Item Analysis; Junior High School Students; *Objective Tests; *Response Style (Tests); Standardized Tests; *Statistical Analysis; Test Construction; Test Reliability; *Test Results, Undergraduate Students
IDENTIFIERS *Answer Changing (Tests)

ABSTRACT

The initial study was based on responses of 103 examinees on an objective classroom test. Results indicated that response changes increased item difficulties as well as biserial and point biserial correlations. Test reliability was not decreased. This implies that not all students benefit from changing initial responses. The personal biserial correlation may be useful in identifying individuals who profit from response changes. The study was replicated on 239 students taking a standardized achievement test. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137411

EFFECTS OF EXAMINEE RESPONSE CHANGES ON ITEM AND TEST CHARACTERISTICS

Linda Crocker

Jeri Benson

University of Florida

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM006 220

Paper presented at the annual meeting of the American Educational Research Association held in New York, N.Y., April, 1977.

Session 9.14

EFFECTS OF EXAMINEE RESPONSE CHANGES ON ITEM AND TEST CHARACTERISTICS[1]

Linda Crocker Jeri Benson[2]

University of Florida

A question commonly asked by examinees and test constructors alike is, "Should the examinee change his responses to objective test items?" In attempts to answer this question, most investigators in this area have focused on how reponse changes affect the total scores of individual examinees (e.g. McMorris and Leonard, 1976; Mueller and Shwedel, 1975; Reiling and Taylor, 1972; Jacobs, 1972; and Bath, 1967). In these studies the examinees' total test score was used as the primary unit of analysis.

Seldom has the problem been approached from the test constructor's point of view. Yet it might be very useful for the test constructor to know: "How do examinee changes affect test quality?" and "Which test and item characteristics are most likely to be affected by examinee response changes?" To answer these questions the researcher must look beyond the examinees' total score to item analysis.

The purpose of this empirical study was to determine the effects of examinee response changes on test and item characteristics for objective examinations. Specifically the following questions were investigated:

(1) How are item difficulties affected by examinee response changes?

(2) How are item statistics, such as biserial and point biserial correlation coefficients affected by examinee response changes?

(3) How is test reliability (i.e. internal consistency) affected by examinee response changes?

(4) How are examinee personal biserial correlations affected by response changes?

---

[1]This study was supported in part by the Institute for Development of Human Resources at the University of Florida.

[2]We are grateful to Mrs. Faye Cake, Director of the Alachua County Teacher Education Center and Mr. William Cliett, Assistant Principal of Fort Clarke Middle School who supported this study and provided data which were used in the analysis.

(5) Does the use of a "Don't Know" option affect examinee item response changes?

(6) What are the characteristics of items which have high rates of response changes?

An important aspect of this study was to have replication across different examinee populations and different types of objective examinations to test the generalizability of the findings.

METHOD

Procedures

Prior to the item analyses performed in this study, tests were administered to examinees using standard machine scorable answer sheets and soft-lead pencils. Examinees received no special instructions about response changes and took the examinations as a normal part of their academic program. The tests were then scored using the examinees' final responses. The tests were re-scored a second time using the examinees' initial response. (A preliminary pilot test had shown that erasures on the answer sheet would be readily detected by visual inspection.) The new answer sheets were prepared by the investigators based upon student erasures on the original answer sheets. In those few cases where the examinee had made more than one answer change per item, one of the erased responses was randomly selected to be coded as the initial response.

An item analysis was conducted on both sets of data to yield item difficulties, biserial correlations between item scores and total scores, point biserial correlations and personal biserial correlations. In addition, estimates of test internal consistency were computed for both sets of data using the Kuder Richardson 20 procedure.

Sample and Instruments

Answer sheets for the first study were obtained from 103 graduate and undergraduate students enrolled in an introductory course in testing and measurement. The 35 item test was a regular unit examination based on course objectives.

To test the generalizability of these results for another student population in a different testing situation, answer sheets from 289 seventh grade students on the 40 item Metropolitan Achievement Mathematics Comprehension Subtest (MAT) were used. These two student populations should have been dissimilar enough in terms of age and test wiseness, and the tests should have differed sufficiently to determine whether results of the study would have widespread generalizability. The MAT also had a "Don't Know" option for each item, which was not used on the classroom test for the college student group.

RESULTS

In general the findings could be summarized as follows:

1. Average item difficulties showed slight positive gains due to examinee response changes for both samples. (See Tables 1 and 2.) However, the group mean gains on total test score were not statistically significant. Despite the small size of the observed increases, it should be noted that p-values increased on 32 out of 35 items for the college examinees and on 39 out of 40 items for the seventh graders.

2. In general item discrimination statistics were relatively unaffected by changes in student responses. For the college examinees (see Table 1) there was little or no shift in the point-biserial correlations between items and total test scores or in the discrimination indexes. For the seventh-graders (see Table 2), the point-biserial values were equally stable. Biserial r values were also

examined for this sample and they too showed little or no effect due to response changes.[3]

3. Internal consistencies of the two tests were relatively unaffected by examinee response changes (Tables 1 and 2), inspite of the fact that the mean number of response changes per item for the college group was 6.9 and for the seventh grade group was 11.6. Thus, test reliability does not appear to be adversely affected when examinees change their answers.

4. The personal biserial index is essentially the biserial correlation used in item analysis, applied to people instead of items (Fischer, 1970). It is the biserial correlation computed across items for a person's item scores (0 or 1) and the proportion of people answering the items correctly.

Personal biserial correlations were calculated for the college examinee group only (see footnote 3). There were no differences in median personal biserial for the college examinees from their first response (r_{perbis} = .37) to their changed response (r_{perbis} = .37). There were no observed differences in the ranges of the personal biserial for the college examinees from their first response (-.13 - .67) to their changed response (-.12 - .67). It was noted that for those examinees who changed only a few answers (1 to 4 changes) personal biserials had a tendency to increase. Of those examinees who made no answer changes the personal biserial was relatively unchanged. The greatest shifts in personal biserials were observed for examinees making many changes in their answers (5 to 11 changes), but directionality of the shifts were not consistent.

5. On 39 out of 40 items, students who originally chose the "Don't Know"

---

[3]Readers should note that the item statistics presented for the college examinees are: difficulty, point-biserial r and index of discrimination. For the seventh graders the item statistics presented are difficulty, point-biserial r and biserial r. This alternation in item statistics reported was necessary because of differences in the answer sheets used at the university and public school levels. Different optical scanning equipment and different item analysis programs had to be used.

option, later changed their responses. The total frequency of "Don't Know" first responses was 344 or an average of 8.6 students per item. The total frequency of "Don't Know" responses after changes was 96, or an average of 2.4 students per item. Thus it is obvious that a major factor in response changes among the seventh-graders was the shift from the "Don't Know" to another option on the test. (Further examination of item response changes revealed that students changed from the "Don't Know" to the correct answer approximately one-third of the time. Since there were 4 possible responses in addition to "Don't Know," it is obvious that students made use of partial knowledge in choosing the correct answer.)

6. To investigate the characteristics of items which had high rates of response change, the 10 items on each test with the greatest number of changes were identified. For the college examinees, these were items with an average of 11 response changes per item; for the seventh graders, these were items with an average of 24 response changes per item. For these items, the following conditions were observed:

- item difficulties (p) were increased slightly
- item discriminations were increased slightly for the college sample
- point biserial correlations were not affected for the college group
- point biserial and biserial correlations were increased slightly for the seventh grade sample

DISCUSSION

In summary, those who construct and administer tests should be heartened by these results, indicating that a moderate amount of response changing has no adverse affect on test quality. If anything, item discriminations may be slightly improved when examinees change responses.

To answer the question often raised by examinees "Should I change my answers?", the best advice seems to be that response changes improve scores more often than

they lower them (albeit to a very slight degree). In this study among the 106 college examinees, 60 (57%) increased their scores and only 9 (8.5%) decreased their scores by changing item responses. In the replication study of 289 seventh grade examinees, 135 examinees (47%) increased their scores while only 7 (or 2%) actually lost points by changing their responses. Looking at all item responses to the test, for the college examinees, 62% of all item response changes yielded the correct response while 18% of all response changes resulted in loss of the correct response. For the seventh graders, 55% of all item response changes resulted in the correct answer and 10% resulted in an incorrect answer. Thus teachers who advise their students against changing responses may actually do their students a disservice.

TABLE 1

Test and Item Statistics Based on Responses of College Examinees
Before and After Response Changes
(N = 106)

Test and Item Characteristics	First Responses	Changed Responses
Mean Item Difficulty($\bar{p}$)	.65	.68
Range of Item Difficulty	.37-.94	.38-.95
Median Item-Test point-biserial r	.36	.39
Range of point-biserial r values	.02-.55	.07-.60
Median Item Discrimination	.39	.40
Range of Item Discrimination	.04-.82	.07-.79
Internal Consistency (KR_{20})	.79	.80
Standard Error of Measurement	2.53	2.51
Overall Test Mean	22.86	23.90
Overall Test Standard Deviation	5.54	5.62

TABLE 2

Test and Item Statistics Based on Responses of 7th Grade Examinees
Before and After Response Changes
(N = 289)

Test and Item Characteristics	First Responses	Changed Responses
Mean Item Difficulty ($\bar{p}$)	.53	.55
Range of Item Difficulty	.27-.93	.29-.93
Median Item-Test point-biserial r	.46	.46
Range of point-biserial r values	.19-.65	.19-.64
Median Item-Test biserial r	.58	.58
Range of biserial r values	.26-.83	.27-.84
Internal Consistency (KR_{20})	.90	.90
Standard Error of Measurement	2.69	2.66
Overall Test Mean	21.42	22.18
Overall Test Standard Deviation	8.68	8.73

REFERENCES

Bath, J. A. Answer-changing behavior on objective examinations. The Journal of Educational Research, 1967, 61, 105-107.

Fischer, F. E. Some properties of the personal biserial index. Journal of Educational Measurement, 1970, 7, 275-277.

Jacobs, S. S. Answer changing on objective tests: Some implications for test validity. Educational and Psychological Measurement, 1972, 32, 1039-1044.

McMorris, R. F. and Leonard, G. Item response changes and cognitive style. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, 1976.

Mueller, D. J. and Shwedel, A. Some correlates of net gain resultant from answer changing on objective achievement test items. Journal of Educational Measurement, 1975, 12, 251-254.

Reiling, E. and Taylor, R. A new approach to the problem of changing initial responses to multiple choice questions. Journal of Educational Measurement, 1972, 9, 67-70.